# Databases and the Hidden Web

## Topics

Lesson 1 – Databases

- Definition
- Attributes
- Examples

Lesson 2 – Grey Literature

- Definition
- Attributes
- Producers
- Examples of search aids, portals and directories, resources

8/17/2012 2

## Objectives

- Students will identify relevant databases and their attributes for open source research.
- Students will recognize grey literature and its value for research. They will review search tools for finding hidden/deep web sources. Students will also review search tools for finding people.

8/17/2012 3

# Lesson 1

## Databases



## What is a Database?

According to SearchSQLServer.com,

- "A database is a collection of information that is organized so that it can easily be accessed, managed, and updated. In one view, databases can be classified according to types of content: bibliographic, full-text, numeric, and images.

- In computing, databases are sometimes classified according to their organizational approach. The most prevalent approach is the relational database, a tabular database in which data is defined so that it can be reorganized and accessed in a number of different ways. A distributed database is one that can be dispersed or replicated among different points in a network. An object-oriented programming database is one that is congruent with the data defined in object classes and subclasses."

SearchSQLServer

LESSON 1



## Databases

Databases provide:

- Consolidated source of related data
- Structured data
- Data in one location and built-in search mechanism

Database attributes:

- Many located on the Internet
- Content is hidden from World Wide Web search engines
- May require registration or a subscription

LESSON 1

## Database Query Techniques

- Boolean Operators – Similar to World Wide Web search engines, but the Boolean is more sophisticated and precise
- Metadata
- Structured Query Language (SQL)
  Example: SDWIS site within Envirofacts of the U.S. EPA at http://oaspub.epa.gov/enviro/sdw_form_v2.create_page?state_abbr=UT

LESSON 1

8/17/2012 7

## Databases - Examples

- **AirNav: www.airnav.com**
- Army Environmental Database: http://aec.army.mil/usaec/reporting/aedb-roo.html
- Army Recruiter/Reserve Center locator: www.goarmy.com/locate-a-recruiter.html
- EUROPAGES: www.europages.com/
- FedBizOpps: www.fbo.gov/?s=main&mode=list&tab=list
- Global Incident Map (subscription only): www.globalincidentmap.com
- **Global Terrorism Database: www.start.umd.edu/gtd/**
- GPO's Federal Digital System (FDsys): www.gpo.gov/fdsys/search/advanced/advsearchpage.action
- Helis.com Database: www.helis.com/database

LESSON 1

8/17/2012 8

## Databases - Examples

- Ingentaconnect: www.ingentaconnect.com/
- National Transportation Safety Board Aviation Accident Database & Synopses: www.ntsb.gov/aviationquery/index.aspx
- National Vulnerability Database: http://nvd.nist.gov/
- OAIster®database: www.oclc.org/oaister
- Plants Database: http://plants.usda.gov/java/
- ThomasNet: www.thomasnet.com/
- **U.S. EPA Water: Applications & Databases http://water.epa.gov/scitech/datait/index.cfm**
- **U.S. National Library of Medicine's databases PubMed: www.ncbi.nlm.nih.gov/pubmed/**

LESSON 1

8/17/2012 9

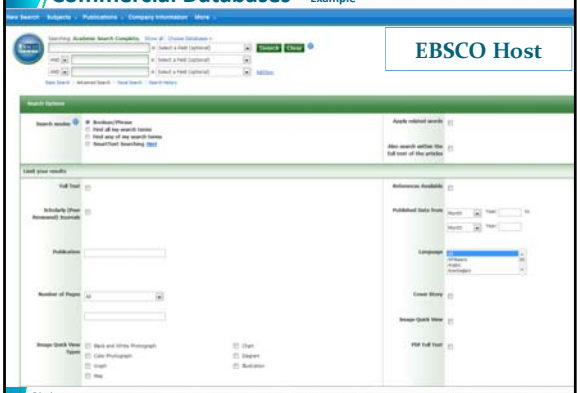
## Commercial Databases

- Authenticated
- Edited
- Value-added features
- Most are fee-based

LESSON 1

8/17/2012 10

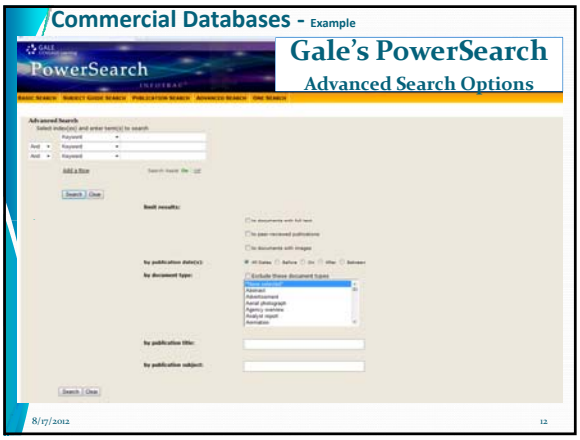
## Commercial Databases - Example



EBSCO Host

8/17/2012 11

## Commercial Databases - Example

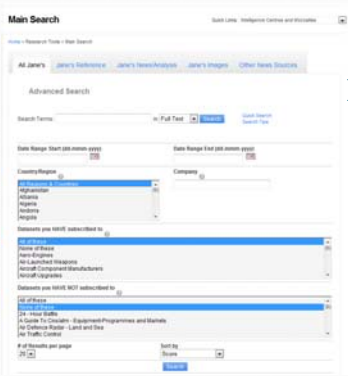



Gale's PowerSearch

Advanced Search Options

8/17/2012 12

## Commercial Databases - Example




**IHS Jane's:**
**Defense & Security Intelligence & Analysis**
**Advanced Search**

8/17/2012 13

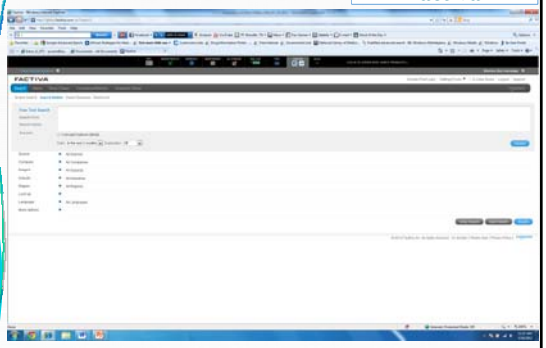
## Commercial Databases - Example

Nexis Power Search

8/17/2012 14

## Commercial Databases - Example

Factiva




8/17/2012 15

## Hands-on Exercise



# Lesson 2

## Grey Literature



## What is Grey Literature?

**Grey Literature**
is also known as
- Gray Literature
- Greylit
- the Hidden Web
- the Invisible Web
- the Deep Web

Grey literature is defined as:

*"Information produced on all levels of government, academia, business and industry in electronic and print formats not controlled by commercial publishing, i.e., where publishing is not the primary activity of the producing body."*

(ICGL Luxembourg definition, 1997 – Expanded in New York, 2004).

LESSON 2

## Why does Grey Literature Exist?

- Surface web crawling limitations
- Surface web is partially accessed
- Mining capabilities
- One dimensional results
- Inconsistent search results
- Document currency
- Accessibility
- Advertising

LESSON 2

8/17/2012 19

## Grey Literature Attributes

- More than half of the content resides in topic-specific databases
- Not published commercially or indexed by major databases
- May be of questionable quality
- Not always published
- Not widely disseminated (not a goal of the publisher)
- Often is not subject to peer review and therefore must be scrutinized accordingly
- May be fugitive – lasts only a short time or changes rapidly



LESSON 2

8/17/2012 20

## Grey Literature Attributes

- Has dynamic content and URLs
- May be password protected, or have a search box
- May be new pages
- May have orphaned web pages
- Robots are not allowed to search the protocol or metatags
- May be too deep in the website; search engines cannot index
- Has non-indexable content
- Search engines may not have the technology to index the non-HTML content
- May have unconventional formats, e.g., brochures may only be in hard copy or posters

LESSON 2

8/17/2012 21

## Grey Literature - Examples

- Theses and dissertations
- Conference proceedings and abstracts, e-prints, preprints
- Newsletters
- Research reports (completed and uncompleted)
- Technical specifications, standards, manuals, and annual reports
- Government documents
- Corporate documents (annual reports, prospectus, maps, product fact sheets)
- Brochures
- Speeches
- Journal archives
- Electronic networked communication
- Blogs, podcasts, video on the web
- Listserv archives
- Digital libraries
- Spatial data
- Institutional repositories
- Wikis, social media
- Working papers

LESSON 2

8/17/2012 22

## Producers of Grey Literature - Examples

- Government departments and agencies (i.e. municipal, provincial, national)
- Conference sites
- Association sites
- Non-profit economic and trade organizations
- Academic and research institutions
- Societies
- Political parties
- Libraries
- Museums
- Archives
- Businesses and corporations
- Individuals

LESSON 2

8/17/2012 23

## Databases, Directories, and Hidden Content

- The good news is that over **90%** of the Deep Web content **is publicly accessible information**.
- Use search engines to find the query site. Surface search engines may find pages but not the content, i.e., content is hidden.
- Use Deep/Hidden web search tools.

LESSON 2

8/17/2012 24

## Searching the Deep Web

- You need a research plan
  - Determine your topic of interest
  - Categorize your topic
  - Decide what type of resource to search
  - Choose your starting point based on your objective

LESSON 2

8/17/2012 25

## Grey Literature Search Aids – Examples

- Complete Planet:
  www.completeplanet.com
- Google Scholar:
  http://scholar.google.com
- INFOMINE:
  http://infomine.ucr.edu/
- TechXtra:
  www.techxtra.ac.uk

LESSON 2

8/17/2012 26

## Grey Literature Portals and Directories - Examples

- **Air War College Military Quick Portal:** **www.au.af.mil/au/awc/awcgate/awc-port.htm**
- American Economists Association, RFE: Resources for Economists on the Internet: http://rfe.org/
- **Applied Math and Science Education Repository (AMSER): http://amser.org./index.php?P=Home**
- **British Library for Development Studies: http://blds.ids.ac.uk/**
- Central and Eastern European Business Directory: www.ceebd.co.uk/ceebd/

LESSON 2

8/17/2012 27

## Grey Literature Portals and Directories - Examples

- CEOExpress: www.ceoexpress.com
- European Business Directory: www.wlw.de/start/wlw_dach/DE/en/
- Legal Research on International Law Issues Using the Internet: www2.lib.uchicago.edu/~llou/forintlaw.html#directories
- MagPortal: www.magportal.com/c/bus/
- **New York Academy of Medicine: www.nyam.org/library**
- **Office of Scientific and Technical Information: www.osti.gov**

LESSON 2

8/17/2012 28

## Grey Literature Portals and Directories - Examples

- **Open Grey: www.opengrey.eu**
- **Scirus: www.scirus.com/srsapp/advanced/index.jsp?q1=www.scirus.com**
- **Thirty Thousand Feet Aviation Directory: www.thirtythousandfeet.com/director.htm**
- University Libraries, Engineering and Physical Sciences Library, The Virtual Technical Reports Center: www.lib.umd.edu/ENGIN/TechReports/Virtual-TechReports.html

LESSON 2

8/17/2012 29

## Grey Literature Portals and Directories - Examples

- **USA.gov: www.usa.gov**
- **Wall Street ExecutiveLibrary: www.executivelibrary.com/**
- **World Health Organization: http://dosei.who.int/uhtbin/cgisirsi/Tue+Jul+10+20:46:21+MEST+2012/0/49**
- **The WWW Virtual Library: http://vlib.org**
- Yahoo! Directory: http://dir.yahoo.com

LESSON 2

8/17/2012 30

## Grey Literature Resources - Examples

- **Amazon:** www.amazon.com
- **Archnet:** www.archnet.org
- Directory of Open Access Journals: www.doaj.org
- Infoplease: www.infoplease.com
- InterDok Corporation (The Meeting Index and Directory of Published Papers): www.interdok.com
- Intute: www.intute.ac.uk

LESSON 2

8/17/2012 31

## Grey Literature Resources - Examples

- **ipL2:** www.ipl2.org
- **Library of Congress:** www.loc.gov
- Migration Information Source: http://migrationinformation.org
- Oregon State University (Institute for Water and Watersheds): www.transboundarywaters.orst.edu
- **Social Science Research Network:** www.ssrn.com/
- **World Research Institute (EarthTrends and Environmental Information):** www.earthtrends.wri.org

LESSON 2

8/17/2012 32

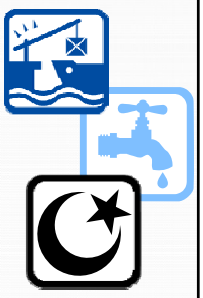
## Hands-on Exercise

## Hands-on Exercise

Question One
- Are there any Panamanian flagged tanker ships named "Columbia"?
- Topic: shipping

Question Two
- Find a report on the quality of drinking water in the city of Mesa, Arizona. Is it good? Is it current?
- Topic: water

Question Three
- According to the most recent census, which state in the United States has the highest number of Muslim Adherents? Which county in Minnesota has the highest number of Muslim Adherents?
- Topic: religion



8/17/2012 34

## Hands-on Exercise

Question Four
- Which resource would provide information on historic earthquakes in Iowa?

Question Five
- Which resource would provide information on Puget Sound's campus disaster plan?

Question Six
- Which resource has the ability to search across all DOE collections?

Question Seven
- Which site provides a chart and map of the percentage of population that is under-nourished?

8/17/2012 35